



Romania's
experience in
using R in the
official
statistics

Organizational
and technical
aspects

Reasons for R @ NIS
Romania
Building R internal
supporting
capabilities
Activities @NIS
Romania

Training and
conferences
Training of
statisticians
Conferences

Romania's experience in using R in the official statistics

Bogdan Oancea; Marian Necula; Ana-Maria Ciuhu;
Ciprian Alexandru; Raluca Dragoescu

National Institute of Statistics, Romania

The 12th International Conference
uRos2024
Piraeus, Greece
November 27-29, 2024

Summary

1 Organizational and technical aspects

Reasons for R @ NIS Romania

Building R internal supporting capabilities

Activities @NIS Romania

2 Training and conferences

Training of statisticians

Conferences

Reasons for NIS Romania

- Cost-related competitive advantage on other software
 - especially migration from **SAS**
- Presenting a **challenge** for each user to actively participate and share knowledge.
- Ongoing **development and updates**, with an expanding collection of packages, printed books, and e-books.
- Our **community**: contributors to R packages and a large network of users.

Building internal supporting capabilities

- Core/starting point:
 - R-omanian Team (voluntary based) working in official statistics and academia
www.r-project.ro
- Departments implementing R in RO-NIS:
 - Demographic and Social statistics
 - Business statistics
 - (Former) Innovative Tools in Statistics Department

Activities @NIS Romania - overview

- Internal projects
 - Web scraping (price index)
 - TEMPO package (access Romanian NIS database)
 - Natural language processing (NLP) for classification (ECOICOP).
 - SAS to R code translation
 - Remote sensing data processing.
 - Sampling procedures in HH surveys: calibration, computing quality indicators
 - Data cleaning, outliers detection, data validation
 - Producing dissemination tables
 - Analysing paradata
 - Small Area Estimation
 - *In progress*: Population Register and New Sampling frame

Activities @NIS Romania

- Web scraping (price index)
 - Web scraping is used to collect price data from a dozen websites.
 - Several R packages are employed, such as rvest, httr(2), RSelenium.
 - Data is stored in a SQLite database.



Activities @NIS Romania

Romania's
experience in
using R in the
official
statistics

Organizational
and technical
aspects

Reasons for R @ NIS
Romania

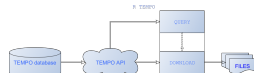
Building R internal
supporting
capabilities

Activities @NIS
Romania

Training and
conferences

Training of
statisticians
Conferences

- R package TEMPO
 - A wrapper package for TEMPO database API.
 - The user can query for the available data.
 - The user can download one or more tables as csv files.



R TEMPO wrapper

Activities @NIS Romania

Romania's
experience in
using R in the
official
statistics

Organizational
and technical
aspects

Reasons for R @ NIS
Romania

Building R internal
supporting
capabilities

Activities @NIS
Romania

Training and
conferences

Training of
statisticians
Conferences

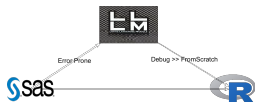
- NLP for ECOICOP classification
 - CPI price data needs to be classified.
 - We tested several word embedding techniques x more than a dozen machine learning algorithms.
 - We obtained results with high accuracy (e.g FastText x Logistic Regression/Random Forest).



2D representation for
Word2Vec SKIP-GRAM

Activities @NIS Romania

- SAS to R code translation
 - There is a growing need to reduce costs and increase statistical production transparency.
 - R programming language and statistical environment is a suitable candidate for porting SAS legacy code.
 - We started porting sampling procedures used in business surveys.¹



SAS to R workflow

¹LLMs show great potential for porting legacy code to FOSS environments, but are error prone.

◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡

Romania's
experience in
using R in the
official
statistics

Organizational
and technical
aspects

Reasons for R @ NIS
Romania

Building R internal
supporting
capabilities

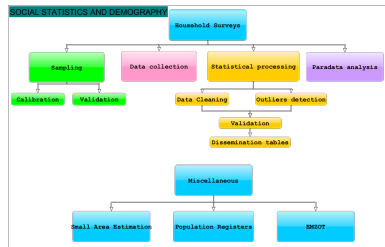
Activities @NIS
Romania

Training and
conferences

Training of
statisticians
Conferences

Activities @NIS Romania

- R for household surveys
 - R programming language and statistical environment is used across the production pipeline of statistics based on household surveys.
 - R packages cover the range from sampling to documentation.
 - R is also used as a backbone for current projects in social statistics and demography.



R for Social Statistics and Demography

Training of statisticians

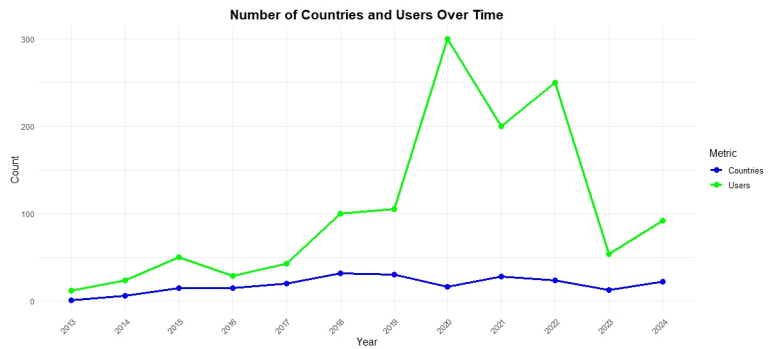
Currently:

- 3-days internal trainings for beginners and intermediate levels

In the past:

- 6 months training-on-the-job sessions (starting 2016)
- 3 to 5 days courses
- one day courses

Insights by uRos





Romania's
experience in
using R in the
official
statistics

Organizational
and technical
aspects

- Reasons for R @ NIS
Romania
- Building R internal
supporting
capabilities
- Activities @NIS
Romania

Training and
conferences

- Training of
statisticians
- Conferences

Thank you!
www.r-project.ro